IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

NPTEL

Smart Grid – Part I

Transforming the Traditional Electrical Grid

Dr. Sudip Misra
Associate Professor
Department of Computer Science and Engineering
IIT KHARAGPUR
Email: smisra@sit.iitkgp.ernet.in
Website: http://www.cse.iitkgp.ac.in/~smisra/

Introduction

- ✓ Advancement of traditional electrical grid
- ✓ Traditional electrical grid
  - ✓ Energy generation is done in centralized power plants
  - ✓ Energy distribution is one directional – from the power plant to the homes or industries.
  - ✓ Monitoring and restoration of grid is done manually
  - ✓ Uni-directional communication
- ✓ Smart Grid –
  - ✓ Achieve high reliability in power systems
  - ✓ A cyber-physical system equipped with sustainable models of energy production, distribution, and usage

What is Smart Grid

- ✓ Smart grid is conceptualized as a planned nationwide network that uses information technology to deliver electricity efficiently, reliably, and securely.
- ✓ Smart grid is also named as –
  - ✓ Electricity with a brain
  - ✓ The energy internet
  - ✓ The electronet
- ✓ According to the definition given by NIST, smart grid is – “*a modernized grid that enables bidirectional flows of energy and uses two-way communication and control capabilities that will lead to an array of new functionalities and applications.*”

Source: https://www.nist.gov/engineering-laboratory/smart-grid/about-smart-grid/smart-grid-beginners-guide

Benefits of Smart Grid

- ✓ Benefits associated with the Smart Grid include:
  - ✓ More efficient transmission of electricity
  - ✓ Quicker restoration of electricity after power disturbances
  - ✓ Reduced operations and management costs for utilities, and ultimately lower power costs for consumers
  - ✓ Reduced peak demand, which will also help lower electricity rates
  - ✓ Increased integration of large-scale renewable energy systems
  - ✓ Better integration of customer-owner power generation systems, including renewable energy systems
  - ✓ Improved security
- ✓ Using smart grid, both the consumers and the energy service providers or stakeholders get benefited.

Benefits of Customers

- ✓ For consumers, the benefit of using smart grid are as follows:
  - ✓ Updated information on their energy usage in real-time
  - ✓ Enabling electric cars, smart appliances, and other smart devices to be charged
  - ✓ Program the smart devices to run during off-peak hours to lower energy bills
  - ✓ Different pricing options

Benefits to Stakeholders

- ✓ For stakeholders, the benefit of using smart grid are as follows:
  - ✓ Increase grid reliability
  - ✓ Reduce the frequency of power blackouts and brownouts
  - ✓ Provide infrastructure for monitoring, analysis, and decision-making
  - ✓ Increase grid resiliency by providing detailed information
  - ✓ Reduce inefficiencies in energy delivery
  - ✓ Integrate the sustainable resources of wind and solar alongside the main grid
  - ✓ Improve management of distributed energy resources, including micro-grid operations and storage management.

Properties of Smart Grid

- ✓ Consumer Participation
  - ✓ Real-time monitoring of consumption
  - ✓ Control of smart appliances
  - ✓ Building Automation
- ✓ Real-time Pricing
- ✓ Distributed Generation
  - ✓ Integration of renewable energy resources
  - ✓ Integration of micro-grid

Properties of Smart Grid (Contd.)

- ✓ Power System Efficiency
  - ✓ Power Monitoring
  - ✓ Asset Management and optimal utilizations
  - ✓ Distribution Automation and Protection

- ✓ Power Quality
  - ✓ Self-Healing
  - ✓ Frequency Monitoring and Control
  - ✓ Load Forecasting
  - ✓ Anticipation of Disturbances

Smart Grid Architecture

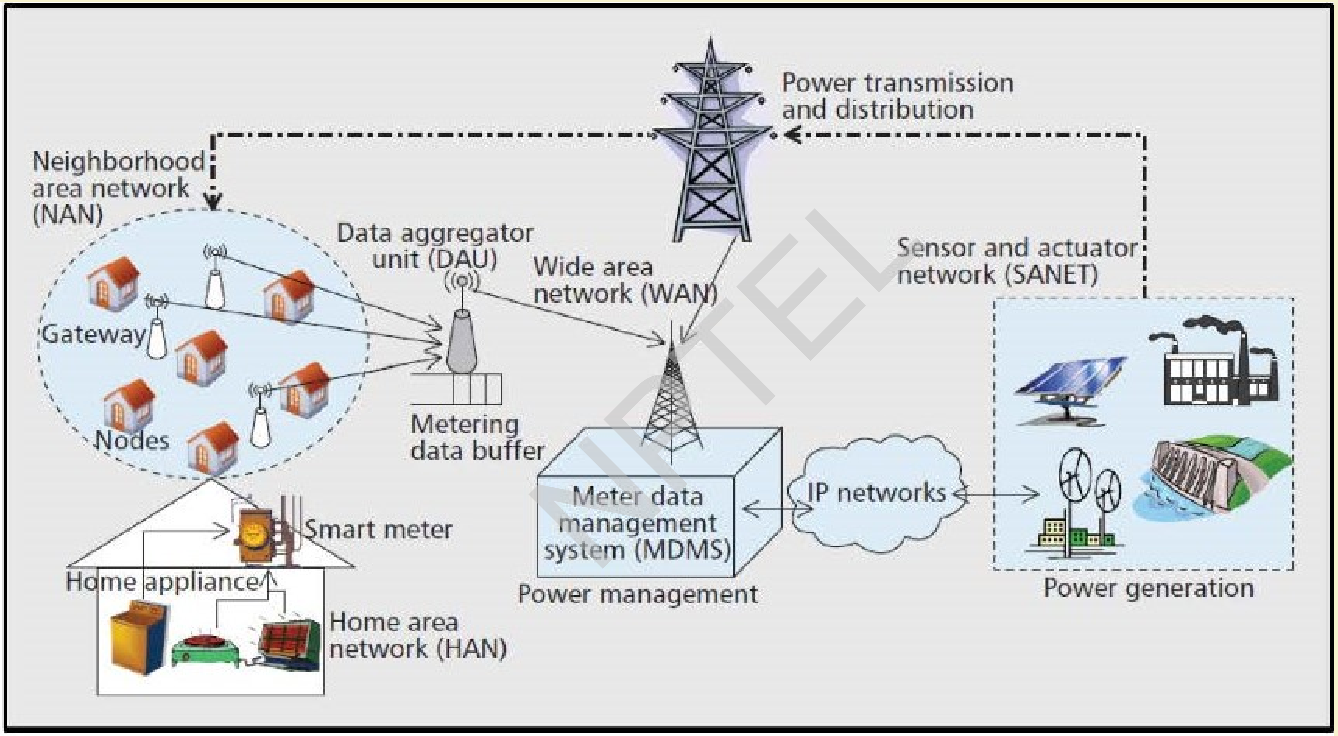
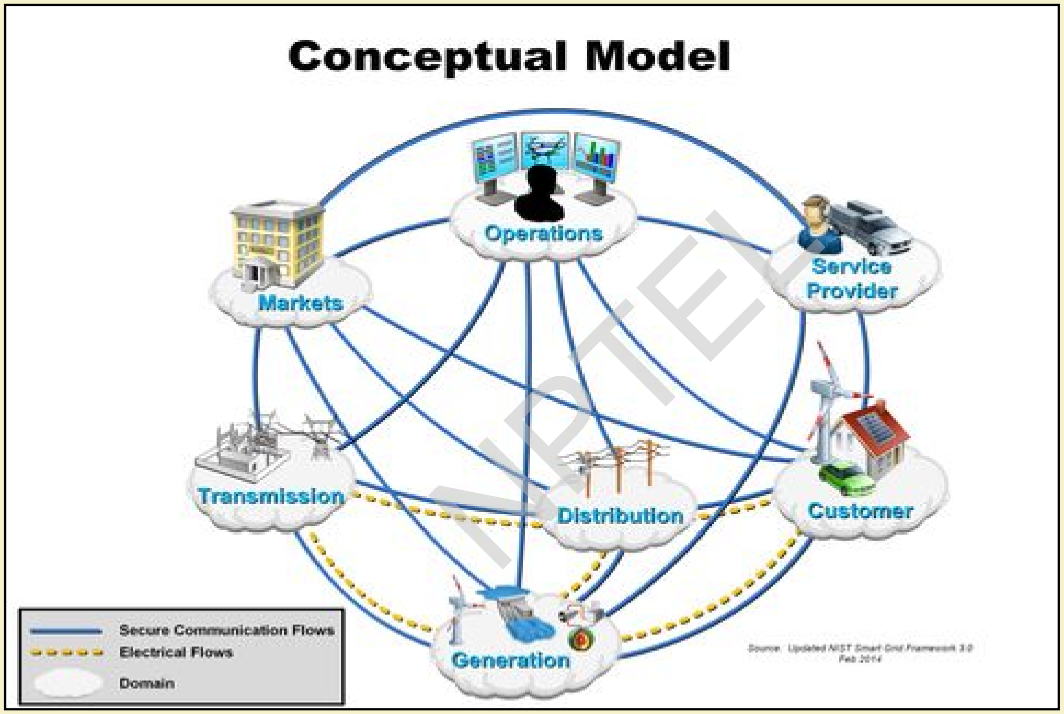


Fig 1: Basic architecture of smart grid [D. Niyato and P. Wang, IEEE CM, 2012]

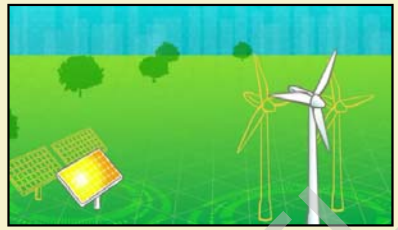
Smart Grid Domains



Source: NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0

Components of Smart Grid

Smart Home



Renewable Energy

Consumer Engagement

Operation Center

Distribution Intelligence

Plug-in Electric Vehicle

Source: https://www.smartgrid.gov/the_smart_grid/

Smart Home

- ✓ Smart home uses emerging smart grid technologies to save energy, seek out the lowest rates, and contribute to the smooth and efficient functioning of our electric grid
- ✓ The interactive relationship between the grid operators, utilities, and consumers helps in proper functioning of smart grid technologies
- ✓ Computerized controls in smart homes helps to minimize energy use at times when the power grid is under stress from high demand, or even to shift some of their power use to times when power is available at a lower cost, i.e., from on-peak hours to off-peak hours

Smart Home (Contd.)

- ✓ Smart home depends on –
  - ✓ Smart meters and home energy management systems
  - ✓ Smart appliances
  - ✓ Home power generation

Smart Home (Contd.)

- ✓ Smart Meters
  - ✓ Provide the Smart Grid interface between consumer and the energy service provider
  - ✓ Operate digitally
  - ✓ Allow for automated and complex transfers of information between consumer-end and the energy service provider
  - ✓ Help to reduce the energy costs of the consumers
  - ✓ Provides information about usage of electricity in different service areas to the energy service providers

Smart Home (Contd.)

- ✓ Home energy management systems
  - ✓ Allows consumers to track energy usage in detail to better save energy
  - ✓ Allows consumers to monitor real-time information and price signals from the energy service provider
  - ✓ Allows to create settings to automatically use power when prices are lowest
  - ✓ Avoids peak demand rates
  - ✓ Helps to balance the energy load in different area
  - ✓ Prevents blackouts
- ✓ In return, the service provider also may choose to provide financial incentives

Smart Home (Contd.)

- ✓ Smart Appliances
  - ✓ Automated and robust in nature
  - ✓ Response to signals from the energy service provider to avoid using energy during times of peak demand
  - ✓ Include consumer controls to override the automated controls
- ✓ By overriding, the consumer can consume energy as per their requirement, while paying minimum is not ensured

Smart Home (Contd.)

- ✓ Home Power Generation
  - ✓ Power generation system at consumers-end
  - ✓ Rooftop solar electric systems
  - ✓ Small wind turbines
  - ✓ Small hydropower System
  - ✓ Home fuel cell systems – produce heat and power from natural gas
- ✓ Surplus energy generated by the home power generation systems can be fed back into the grid
- ✓ In case of "Islanding", a home can have power from distributed resources, i.e., home power generation systems

Renewable Energy

- ✓ According to the International Energy Agency –
  - ✓ *"Renewable energy is derived from natural processes that are replenished constantly. In its various forms, it derives directly from the sun, or from heat generated deep within the earth. Included in the definition is electricity and heat generated from solar, wind, ocean, hydropower, biomass, geothermal resources, and biofuels and hydrogen derived from renewable resources."*
- ✓ Reduced environmental pollution
- ✓ Consumers capable of generating energy from renewable energy resources are less dependent on the micro-grid or main grid
- ✓ In addition to that, they can supply surplus amount of energy from the renewable resources and can make profit out of it

Consumer Engagement

- ✓ Consumers can –
  - ✓ Save energy with proper scheduling of smart home appliances
  - ✓ Pay less for consuming energy in off-peak hours
- ✓ Energy service provider gives incentives based on the energy consumption of the consumer and they can save money
- ✓ Consumers' involvement in following ways:
  - ✓ Time-of-Use pricing
  - ✓ Net metering
  - ✓ Financial incentives

Consumer Engagement (Contd.)

- ✓ In Time-of-Use pricing
  - ✓ The consumers are encouraged to consume energy in off-peak hours when the energy load is less
  - ✓ Throughout the day, the energy load on the grids are dynamic
- ✓ In on-peak hours, if the requested amount of energy is higher, it leads to –
  - ✓ Less-efficient energy distribution
  - ✓ More pollution – it depends on the non-renewable energy resource to meet the peak requirement
- ✓ Home energy management system tries to schedule the smart appliances in off-peak hours
  - ✓ To ensure efficient service
  - ✓ To pay less

Consumer Engagement (Contd.)

- ✓ Net metering
  - ✓ It is feasible with the installation of smart meters
  - ✓ Consumers are paid high, if they are supplying excess amount of generated energy to the grid in on-peak hours
  - ✓ The price is less in case of off-peak hours

- ✓ Final bills to be paid by the consumers depends on
  - ✓ The in-flow of energy (from the grid to the consumers-end)
  - ✓ The out-flow of energy (from the consumers-end to the grid)

- ✓ The consumer may get incentives from the energy service provider at the end of the year based on the net metering value

Consumer Engagement (Contd.)

- ✓ Financial Incentives
  - ✓ Energy service provider offers some financial incentives for the consumers' participation
  - ✓ Incentives for shifting operation of appliances to the off-peak hours
  - ✓ Incentives for using stored energy at the battery installed at the consumers-end or at the plug-in hybrid electric vehicles (PHEVs)
- ✓ Smart grid enables consumers engagement to a large extend
- ✓ Consumers get financial incentives by different means from the energy service providers
- ✓ Energy service providers maintain efficient and load balancing energy distribution

Thank You!!

Smart Grid – Part II

Transforming the Traditional Electrical Grid

Dr. Sudip Misra
Associate Professor
Department of Computer Science and Engineering
IIT KHARAGPUR
Email: smisra@sit.iitkgp.ernet.in
Website: http://www.cse.iitkgp.ac.in/~smisra/

Operation Centers

- ✓ Drawbacks of traditional operation centers
  - ✓ Tries to make sure the amount of generated energy is getting used
  - ✓ The grid is unstable, if the grid voltage drops due to excess energy generation
  - ✓ Limited control capabilities
  - ✓ No means to detect oscillation which leads to blackout
  - ✓ Limited information about the energy flow through the grid
- ✓ Smart grid
  - ✓ Provides information and control on the transmission system
  - ✓ Makes the energy grid more reliable
  - ✓ Minimize the possibility of widespread blackouts

Operation Centers (Contd.)

- ✓ For monitoring and controlling the transmission System in smart grid, phasor measurement unit (PMU) is used
- ✓ PMU samples voltage and current with a fixed sample rate at the installed location
- ✓ It provides a snapshot of the active power system at that location
- ✓ By increasing the sampling rate, PMU provides the dynamic scenario of the energy distribution system
- ✓ PMU helps to identify the possibility of blackout in advance
- ✓ Multiple PMUs form a phasor network
- ✓ Collected information by the phasor network is analyzed at centralized system, i.e., Supervisory Control And Data Acquisition (SCADA) system

Operation Centers (Contd.)

- ✓ Self-healing of grid
  - ✓ Dampen unwanted power oscillations
  - ✓ Avoid unwanted flows of current through the grid
  - ✓ Reroute power flows in order to avoid overloading in a transmission line
  - ✓ This is part of distribution intelligence
- ✓ Demand side energy distribution
  - ✓ Energy supply is done based on the requirement of the consumers
  - ✓ The consumers pay according the consumed energy and price decide by the energy service provider at that time
- ✓ In smart grid, the energy distributors can form coalition and serve the energy requirement in a specific geographic location

Distribution Intelligence

- ✓ Distribution intelligence <u>means the energy distribution systems equipped with smart IoT devices</u>
- ✓ Along with smart meters, distribution intelligence can –
  - ✓ Identify the source of a power outage
  - ✓ Ensure power flow automatically by combining automated switching
  - ✓ Optimize the balance between real and reactive power
- ✓ Reactive power:
  - ✓ Devices that store and release energy
  - ✓ Cause increased electrical currents without consuming real power
- ✓ Intelligent distribution System
  - ✓ Maintains the proper level of reactive power in the System
  - ✓ Protect and control the feeder lines

Plug-In Electric Vehicles

- ✓ Smart Grids have the infrastructure needed to enable the efficient use of plug-in electric vehicle (PEVs)
- ✓ Using PEVs –
  - ✓ Reduce dependency on oil
  - ✓ No pollution when running on electricity
- ✓ PEVs rely on power plants to charge their batteries
- ✓ Energy service provider encourages the consumers to charge batteries of PEVs in off-peak hours
- ✓ PEVs also can be used as an energy source in on-peak hours
- ✓ PEVs get incentives from energy service provider for providing energy to the grid through discharging

Smart Grid Communication

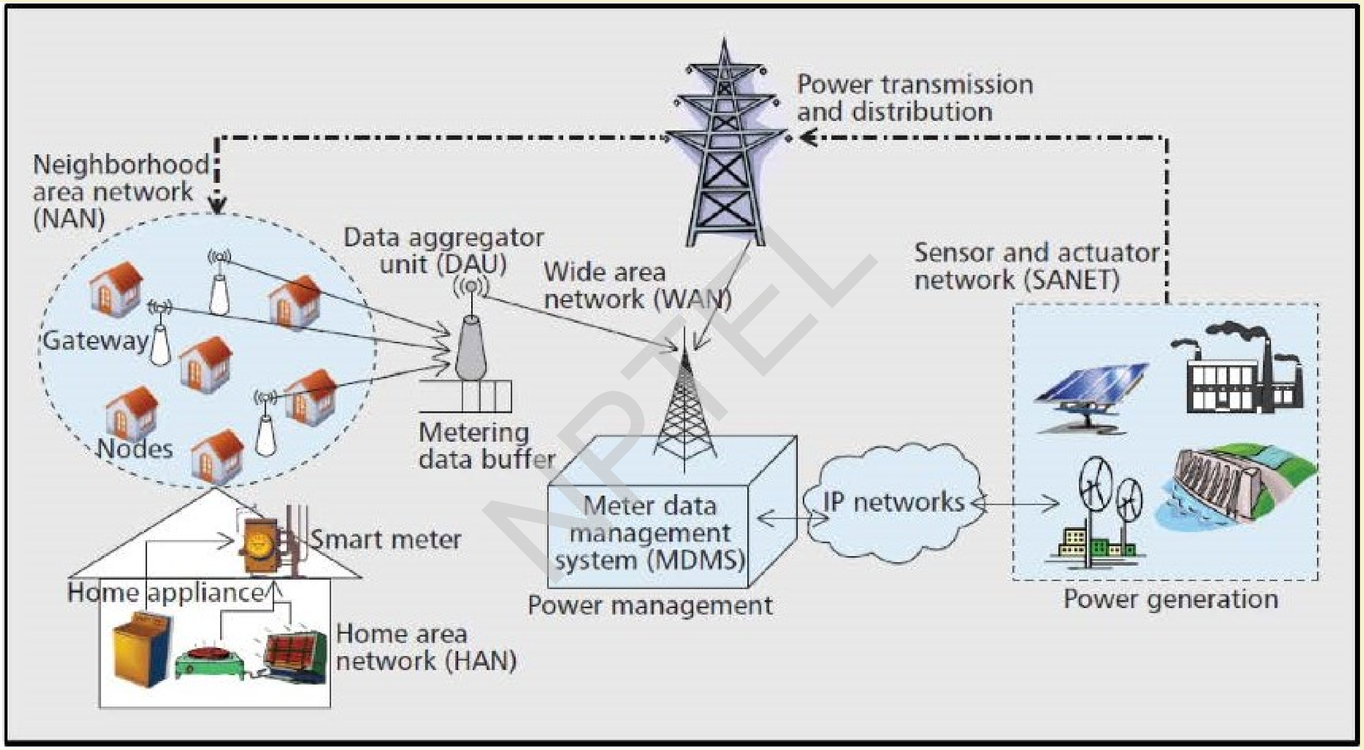


Fig 2: Smart Grid Communication[D. Niyato and P. Wang, IEEE CM, 2012]

Smart Grid Communication (Contd.)

- ✓ Components for smart grid communication are as follows:
  - ✓ Smart Home Appliances
  - ✓ Smart Meters
  - ✓ Gateways
  - ✓ Data Aggregator Units (DAUs)
  - ✓ Meter Data Management Systems (MDMSs)
- ✓ Different networks associated with smart grid communication
  - ✓ Home Area Networks (HANs)
  - ✓ Neighborhood Area Networks (NANs)
  - ✓ Wide Area Networks (WANs)
  - ✓ IP Networks
  - ✓ Sensors and Actuators Networks (SANETs)

Smart Grid Communication (Contd.)

- ✓ For Smart Home Appliances, the available protocol are as follows:
- ✓ C-Bus:
  - ✓ Data Rate: 3500 bits/sec
  - ✓ Able to handle cable lengths upto 1000 m
- ✓ DECT
  - ✓ Data rate: 64000 bits/sec
  - ✓ Operates in 1880 – 1930 MHz
- ✓ EnOcean
  - ✓ Data rate: 9600 bits/sec
  - ✓ Operates in 902 MHz in North America
- ✓ Universal Power line Bus
  - ✓ Data rate: 480 bits/sec
  - ✓ Enable two-way communication protocol

Smart Grid Communication (Contd.)

- ✓ Thread
  - ✓ Data Rate: 20-250 Kbits/sec
  - ✓ IPv6 addressing based 6LowPAN networking protocol
- ✓ Zigbee
  - ✓ Data Rate: 20-250 Kbits/sec
  - ✓ Operates in 2.4 GHz band
  - ✓ IEEE 802.15.4 protocol
  - ✓ Communication range ~100 m
- ✓ Simplified Cable Solution (SCS)
  - ✓ Data rate: 9.6 Kbits/sec
  - ✓ Works on twisted pair
  - ✓ Developed based on OpenWebNet

Smart Grid Communication (Contd.)

- ✓ Smart Meters and Gateways
  - ✓ Each gateway connects few closely located smart meters
  - ✓ Gateways communicate mostly based on WiFi, i.e., IEEE 802.11
  - ✓ Gateways helps in two-way communication
- ✓ Smart meters
  - ✓ Forward the energy consumption information fro the home appliances to the gateways
  - ✓ Forward the billing amount and the control information from the gateways to the home appliances
- ✓ Gateway acts as link between the smart meters and the data aggregator units (DAUs)

Smart Grid Communication (Contd.)

- ✓ Data Aggregator Units (DAUs)
  - ✓ Aggregate the energy consumption or energy request of certain geographical area
  - ✓ Forward the energy consumption information to the centralized coordinator – meter data management system (MDMS)
  - ✓ Maintains a buffer to queue the energy consumption information of the consumers

Smart Grid Communication (Contd.)

- ✓ Meter Data Management Systems (MDMSs)
  - ✓ Act as the centralized coordinator for smart grid communication
  - ✓ Handled by the energy service providers
  - ✓ Part of operation center
  - ✓ Decide the price per unit energy to be paid by the consumers

Smart Grid Security

- ✓ Smart grid is a cyber physical system
- ✓ Following vulnerabilities are there in smart grid
  - ✓ Integrity – credibility of the data collected and transferred over the grid
  - ✓ Availability – accessibility to every grid component as well as to the information transmitted and collected
  - ✓ Dynamic system attacks – based on the previous information same type of request can be replicated by the attacker
  - ✓ Physical threats – physical attack to the smart grid components
  - ✓ Coordinated attacks – cascading failure of systems in smart grid

Smart Grid Security (Contd.)

- ✓ Integrity
  - ✓ Data injection attacks (DIAs)
    - ✓ Manipulation of exchanged data such as sensor readings, feedback control signals, and electricity price signals
    - ✓ Performed by compromising the hardware components (as in the case of Stuxnet), or intercepting the communication links
  - ✓ System Damage
    - ✓ An attacker can manipulate system measurements so that a congested transmission line falsely seems to not have reached its thermal transmission limit
    - ✓ Induce large fluctuations in system dynamics that can lead to tripping additional lines, disconnecting generators, load shedding, or even a system blackout

Smart Grid Security (Contd.)

- ✓ Integrity
  - ✓ Financial benefit
    - ✓ Manipulating the electricity prices
    - ✓ Doing this one can buy energy with lesser price from a service provider and make high profit
  - ✓ Time synchronization attacks
    - ✓ An adversary can manipulate the time reference of the time stamped measured phasors to create a false visualization of the actual system conditions thus yielding inaccurate control and protection actions
    - ✓ Attacks that target PMU time synchronization are known as time synchronization attacks (TSAs)

Smart Grid Security (Contd.)

- ✓ Availability
  - ✓ Accessibility unavailable to every grid component as well as to the information transmitted and collected, whenever needed
  - ✓ Attacks compromising this availability are known as denial of service (DoS) attacks
  - ✓ Block key signals to compromise the stability of the grid and observability of its states
  - ✓ Manipulating generation-load balance

Smart Grid Security (Contd.)

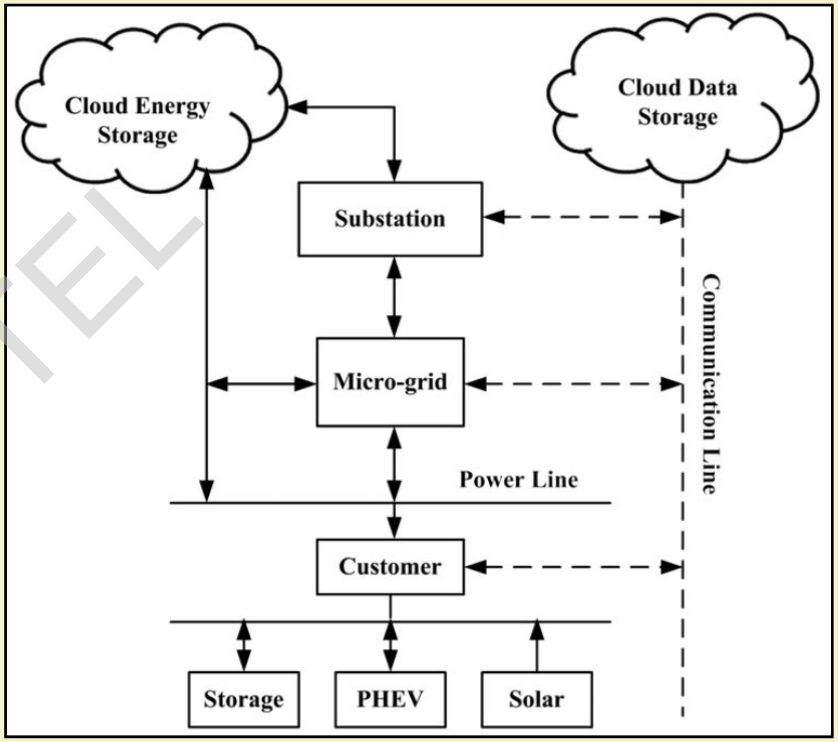
- ✓ Dynamic System Attacks
  - ✓ Replay attacks (RAs)
    - ✓ Injects input data in the system without causing changes to the measurable outputs
  - ✓ In RAs –
    - ✓ Compromises sensors, monitors their outputs
    - ✓ Learns the outputs and repeats them while injecting its attack signal
  - ✓ Dynamic data injection attacks (D-DIA)
    - ✓ Uses knowledge of the grid's dynamic model to inject data that causes unobservability of unstable poles
    - ✓ Can lead to a system collapse
  - ✓ Covert attack
    - ✓ Closed loop version of replay attacks

Smart Grid Security (Contd.)

- ✓ Physical Threats
  - ✓ Attacks a physical component such as a generator, substation, or transmission line is prominent
  - ✓ Physical manipulation of smart meters for energy theft purposes
- ✓ Coordinated Attacks
  - ✓ Power system typically incorporates robustness measures
  - ✓ An attack leading to the failure of one or few components
  - ✓ Exploit the dense interconnections between grid components to launch simultaneous attacks of different types targeting various components

Smart Grid and Cloud Applications

- ✓ In smart grid, cloud applications take a lead in several aspects
  - ✓ Energy management
  - ✓ Information management
  - ✓ Security



S. Bera, S. Misra, and J. J. P. C. Rodrigues, "Cloud Computing Applications for Smart Grid: A Survey," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 5, pp. 1477–1494, May 2015.

Energy Management and Cloud Application

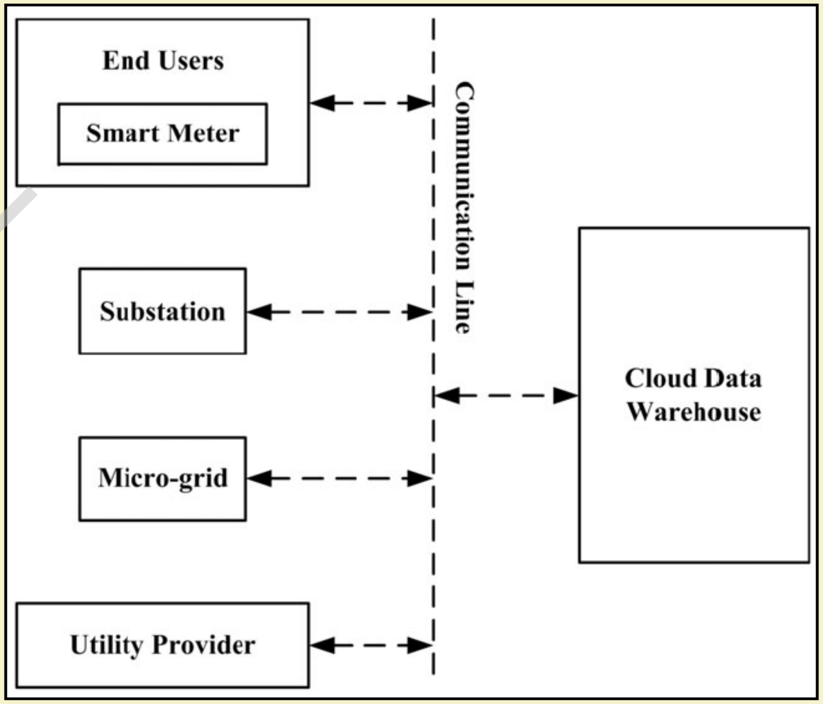
- ✓ The energy management in smart grid can be more efficient by using cloud applications
  - ✓ Cloud-Based Demand Response for fast response times in large scale deployment
  - ✓ Two cloud-based demand response models are proposed as follows:
    - ✓ Data-centric communication and
    - ✓ Topic-based group communication
- ✓ With the integration of cloud, requests from customers are scheduled which are to be executed depending on the available resources, priority, and other applicable constraints
- ✓ Incoming jobs from users are scheduled according to their priority, available resources, and applicable constraints

Energy Management and Cloud Application (Contd.)

- ✓ Integrating cloud computing applications for micro-grid management in the form of different modules such as infrastructure, power management, and service
- ✓ The number of supported customers increases
- ✓ With cloud application, integrate and analyze information streaming from multiple smart meters simultaneously can be done, in order to balance the real-time demand and supply curves
- ✓ Real-time energy usage and pricing information can be shared
- ✓ Mobile agent can be used to monitor power system using cloud computing platform due to the smart grid's heterogeneous architecture

Information Management and Cloud Application

- ✓ Information processing in smart grid fit well with the computing and storage mechanisms available for cloud applications
- ✓ Information from different components, and the supply and demand state conditions can be shared with the help of cloud computing
- ✓ Real-time distributed data management and parallel processing of information can be utilized using smart grid data cloud application

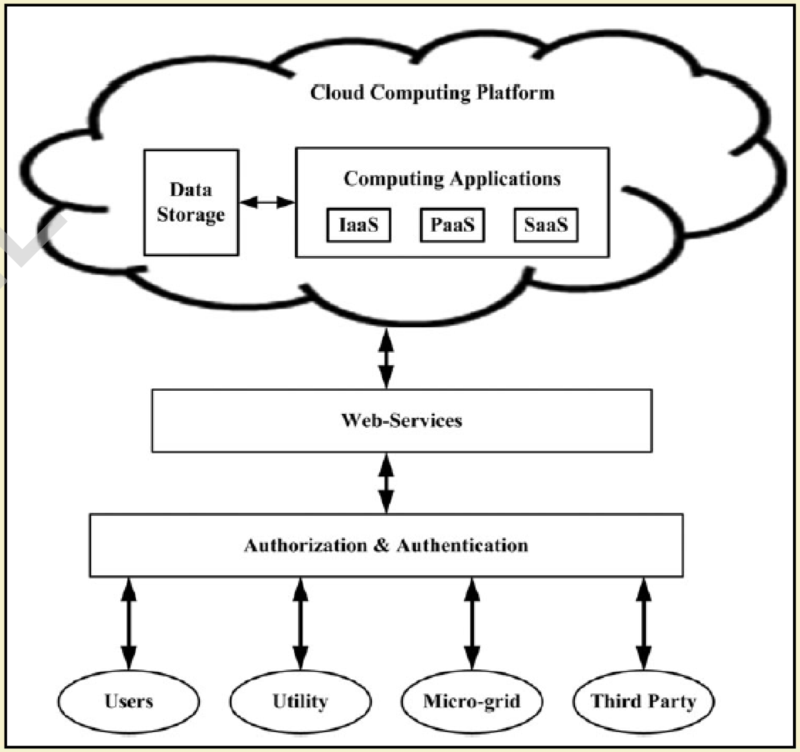


Information Management and Cloud Application (Contd.)

- ✓ With the flexibility of cloud computing, information is retrieved from the data cloud more conveniently in smart grid
- ✓ Dynamic pricing mechanism in smart grid is feasible with the use of cloud application
- ✓ Cloud computing services are used as a dynamic data centers to store the real-time information from the smart meters
- ✓ Use of multi-mobile agent combined with cloud computing for profitable smart grid operation
- ✓ Interactive cooperation using cloud services to support multiple customers and multiple energy sources for large-scale development of smart grid for energy management

Security in Smart Grid and Cloud Application

- ✓ An electric power information security and protection system can be developed using based on cloud security
- ✓ Private cloud platforms are suitable for scaling out and processing millions of data from users
- ✓ Using the cloud computing platform, the electrical utilities can quickly and effectively deal with malicious software



Security in Smart Grid and Cloud Application (Contd.)

- ✓ Security and protection system for electrical power
  - ✓ Servers act as cloud and take decision according to the clients' data
- ✓ Privacy issue in smart grid
  - ✓ Quickly and effectively deal with malicious software with the implementation of cloud computing applications
- ✓ Data storage security for distributed verification in smart grid using cloud application
- ✓ Real-time data can be analyzed and estimated using cloud in smart grid
- ✓ Cloud-based information privacy scheme can be used for smart grid data privacy

References

- ✓ G. Strbac, “Demand side management: Benefits and challenges,” Energy Policy, vol. 36, no. 12, pp. 4419–4426, 2008.
- ✓ M. Such and C. Hill, “Battery energy storage and wind energy integrated into the Smart Grid,” in Proceedings of IEEE PES on Innovative Smart Grid Technologies, Washington, Jan 2012, pp. 1–4.
- ✓ S. Misra, P. V. Krishna, V. Saritha, and M. S. Obaidat, “Learning Automata as a Utility for Power Management in Smart Grids,” IEEE Communications Magazine, vol. 51, no. 1, pp. 98–104, 2013.
- ✓ V. Bakker, M. G. C. Bosman, A. Molderink, J. L. Hurink, and G. J. M. Smit, “Demand Side Load Management Using a Three Step Optimization Methodology,” in Proceedings of the 1st IEEE International Conference on Smart Grid Communications, Gaithersburg, Oct 2010, pp. 431–436.
- ✓ S. Misra, S. Bera, and T. Ojha, “D2P: Distributed Dynamic Pricing Policy in Smart Grid for PHEVs Management,” IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 3, pp. 702–712, Mar 2015.
- ✓ S. Bera, S. Misra, and J. J. P. C. Rodrigues, “Cloud Computing Applications for Smart Grid: A Survey,” IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 5, pp. 1477–1494, May 2015.
- ✓ S. Misra, A. Mondal, S. Banik, M. Khatua, S. Bera, and M. S. Obaidat, “Residential Energy Management in Smart Grid: A Markov Decision Process-Based Approach,” in IEEE International Conference on Internet of Things, Beijing, Chaina, Aug 2013, pp. 1152–1157.
- ✓ A. Mondal and S. Misra, “Game-Theoretic Green Electric Vehicle Energy Networks Management in Smart Grid,” in IEEE International Conference on Advanced Networks and Telecommunications Systems,Dec 2015, pp. 1–6.

References (Contd.)

- A. Molderink, V. Bakker, M. G. C. Bosman, J. L. Hurink, and G. J. M. Smit, “Management and Control of Domestic Smart Grid Technology,” IEEE Transactions on Smart Grid, vol. 1, no. 2, pp. 109–119, Aug 2010.
- M. Erol-Kantarci and H. T. Mouftah, “TOU-Aware Energy Management and Wireless Sensor Networks for Reducing Peak Load in Smart Grids,” in the 72nd IEEE Vehicular Technology Conference Fall, Ottawa, ON, Sept 2010, pp. 1 – 5.
- A. Mondal and S. Misra, “Dynamic Coalition Formation in a Smart Grid: A Game Theoretic Approach,” in Proceedings of IEEE International Workshop on Smart Communication Protocols and Algorithms in conjunction with IEEE ICC, Budapest, Hungary, Jun 2013, pp. 1067 – 1071.
- F. Farzan, F. Farzan, M. A. Jafari, and J. Gong, “Integration of Demand Dynamics and Investment Decisions on Distributed Energy Resources,” IEEE Transactions on Smart Grid, vol. 7, no. 4, pp. 1886–1895, Jul 2016.
- A. Mondal and S. Misra, “Game-Theoretic Energy Trading Network Topology Control for Electric Vehicles in Mobile Smart Grid,” IET Networks, vol. 4, no. 4, pp. 220–228, 2015.
- F. Kamyab, M. Amini, S. Sheykhha, M. Hasanpour, and M. M. Jalali, “Demand Response Program in Smart Grid Using Supply Function Bidding Mechanism,” IEEE Transactions on Smart Grid, vol. 7, no. 3, pp. 1277–1284, May 2016.
- A. Sanjab, W. Saad, I. Guvenc, A. Sarwat, and S. Biswas, "Smart Grid Security: Threats, Challenges, and Solutions," arXiv preprint arXiv:1606.06992 (2016).
- A. Mondal and S. Misra, “Dynamic Data Aggregator Unit Selection in Smart Grid: An Evolutionary Game Theoretic Approach,” in IEEE India Conference, Dec 2014, pp. 1–6.

References (Contd.)

- ✓ P. Samadi, V. W. S. Wong, and R. Schober, "Load Scheduling and Power Trading in Systems With High Penetration of Renewable Energy Resources," IEEE Transactions on Smart Grid, vol. 7, no. 4, pp. 1802–1812, Jul 2016.
- ✓ A. Mondal and S. Misra, "Game-Theoretic Distributed Virtual Energy Cloud Topology Control for Mobile Smart Grid," in IEEE 6th International Conference on Cloud Computing Technology and Science (CloudCom), Dec 2014, pp. 54–61.
- ✓ A. Mondal, S. Misra, and M. S. Obaidat, "Distributed Home Energy Management System With Storage in Smart Grid Using Game Theory," IEEE Systems Journal, pp. 1–10, 2015.
- ✓ C. P. Mediwaththe, E. R. Stephens, D. B. Smith, and A. Mahanti, "A Dynamic Game for Electricity Load Management in Neighborhood Area Networks," IEEE Transactions on Smart Grid, vol. 7, no. 3, pp. 1329–1336, May 2016.
- ✓ X. Liang, X. Li, R. Lu, X. Lin, and X. Shen, "UDP: Usage-Based Dynamic Pricing With Privacy Preservation for Smart Grid," IEEE Transactions on Smart Grid, vol. 4, no. 1, pp. 141–150, Mar 2013.
- ✓ S. Shivshankar and A. Jamalipour, "An Evolutionary Game Theory-Based Approach to Cooperation in VANETs Under Different Network Conditions," IEEE Transactions on Vehicular Technology, vol. 64, no. 5, pp. 2015–2022, May 2015.
- ✓ P. Samadi, H. Mohsenian-Rad, R. Schober, and V. W. S. Wong, "Advanced Demand Side Management for the Future Smart Grid UsingMechanism Design," IEEE Transactions on Smart Grid, vol. 3, no. 3, pp. 1170–1180, Sept 2012.

Thank You!!

IIT KHARAGPUR

NPTEL ONLINE CERTIFICATION COURSES

NPTEL

IIoT: Industrial Internet of Things – Part I

Dr. Sudip Misra

Associate Professor

Department of Computer Science and Engineering

IIT Kharagpur

Email: smisra@sit.iitkgp.ernet.in

Website: http://cse.iitkgp.ac.in/~smisra/

“IoT as a concept has crossed the chasm from slideware to reality with many industries implementing IoT solutions.”

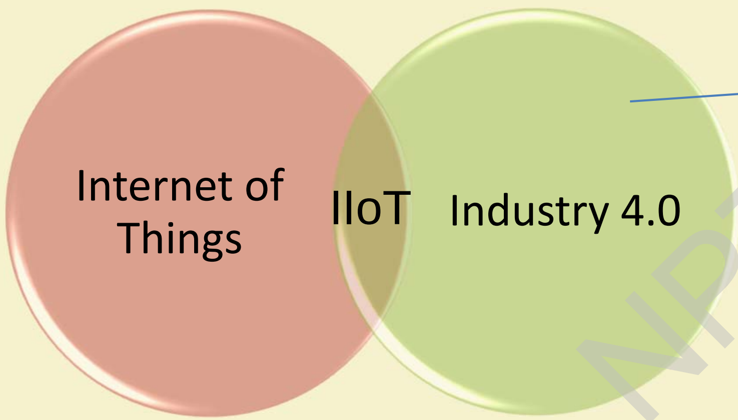
- Paul Howarth, Senior Manager, Corporate Development, CISCO

Source : http://www.mcrockcapital.com

Introduction

- ✓ The main aim of Internet of Things (IoT) is
  - ✓ to globally connect smart 'things' or 'objects' .
  - ✓ objects are uniquely identified.
  - ✓ interoperability among the objects.
- ✓ The Industrial Internet of Things (IIoT) is an application of IoT in industries to modify the various existing industrial systems. IIoT links the automation system with enterprise, planning and product lifecycle.

Introduction (contd.)



- Automation and data exchange in manufacturing technologies
- Cyber-physical systems, the Internet of things and cloud computing
- Smart factory

Fig 1(a) : IIoT as an intersection of industries and IoT

Introduction (contd.)

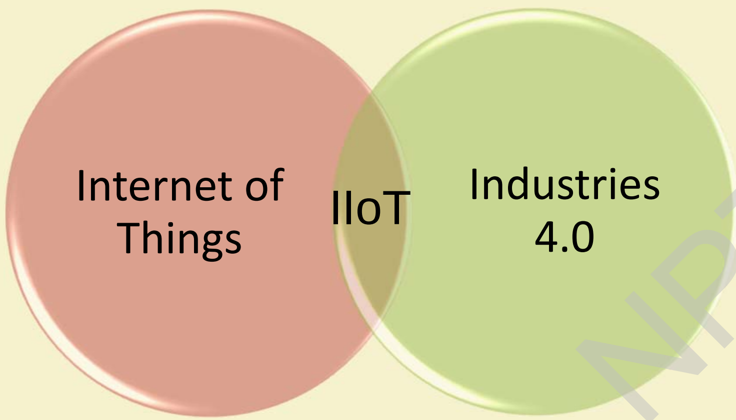


Fig 1(a) : IIoT as an intersection of industries and IoT

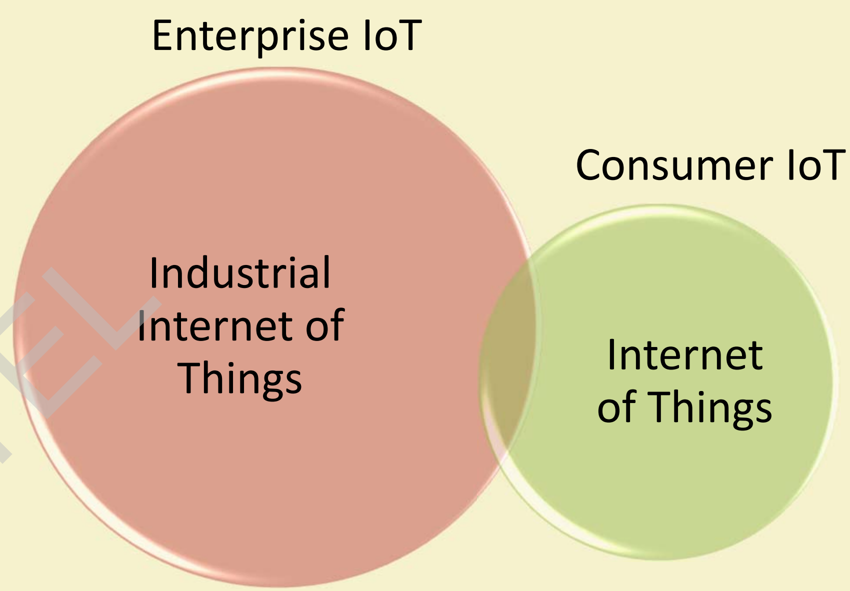


Fig 1(b) : IIoT ≠ IoT

Fig 1 : IIoT Platform

Introduction (contd.)

- ✓ IIoT includes –
  - ✓ machine learning
  - ✓ big data technology
  - ✓ machine - to - machine interaction (M-2-M)
  - ✓ automation.
- ✓ IIoT is supported by huge amount of data collected from sensors. It is based on “wrap & re-use” approach, rather than “rip & replace” approach. (Source : http://www.mhi.org)

Introduction (contd.)

- ✓ 1st Industrial Revolution : Mechanized production
- ✓ 2nd Industrial Revolution : Mass production
- ✓ 3rd Industrial Revolution : Internet evolution and automation
- ✓ 4th Industrial Revolution : IIoT

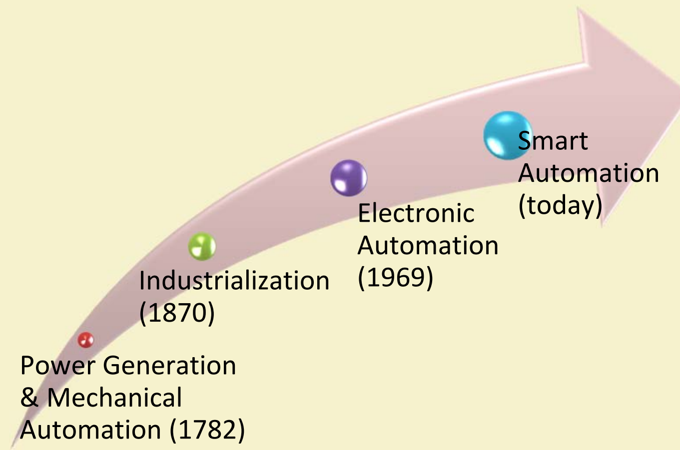


Fig 2: Industry 4.0

Source: http://www.industry40wood.com

Introduction (contd.)

Source: https://www.artika.info

Source: http://www.rehm-group.com

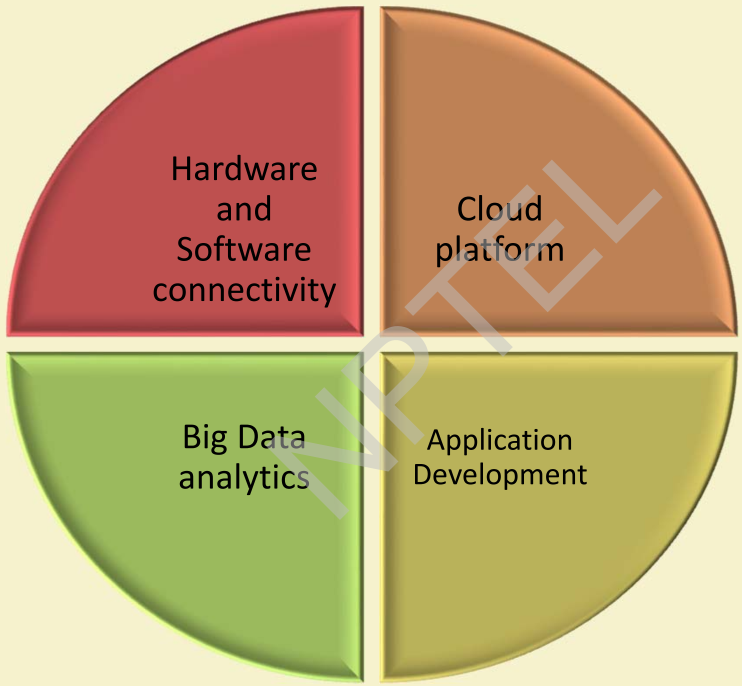
Cloud computing

IIoT : 2nd generation of Internet evolution and 4th Industrial Automation

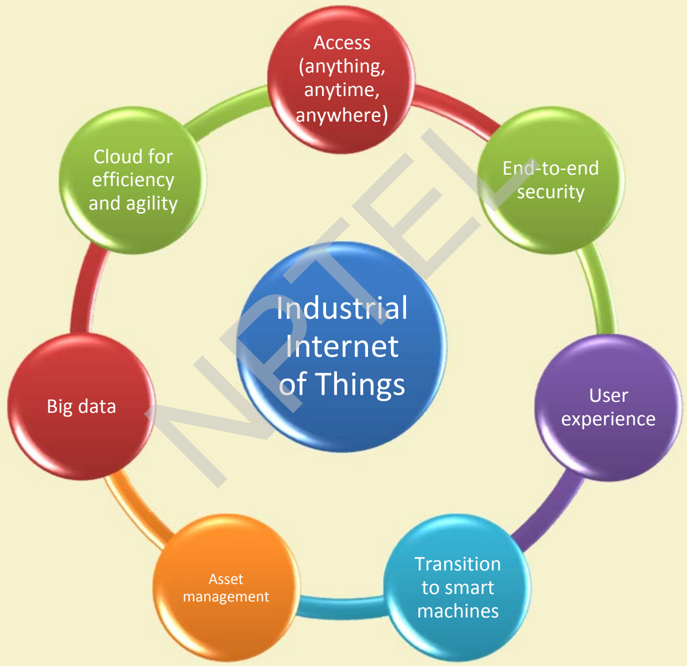
Introduction (contd.)

- ✓ IIoT is a network of
  - ✓ physical objects
  - ✓ systems
  - ✓ platforms
  - ✓ applications
- ✓ These networks can communicate with each other, external environment and other people.
- ✓ The acquisition of IIoT has led to availability and affordability of sensors, processors, and other technologies which facilitates capture and access to real-time information

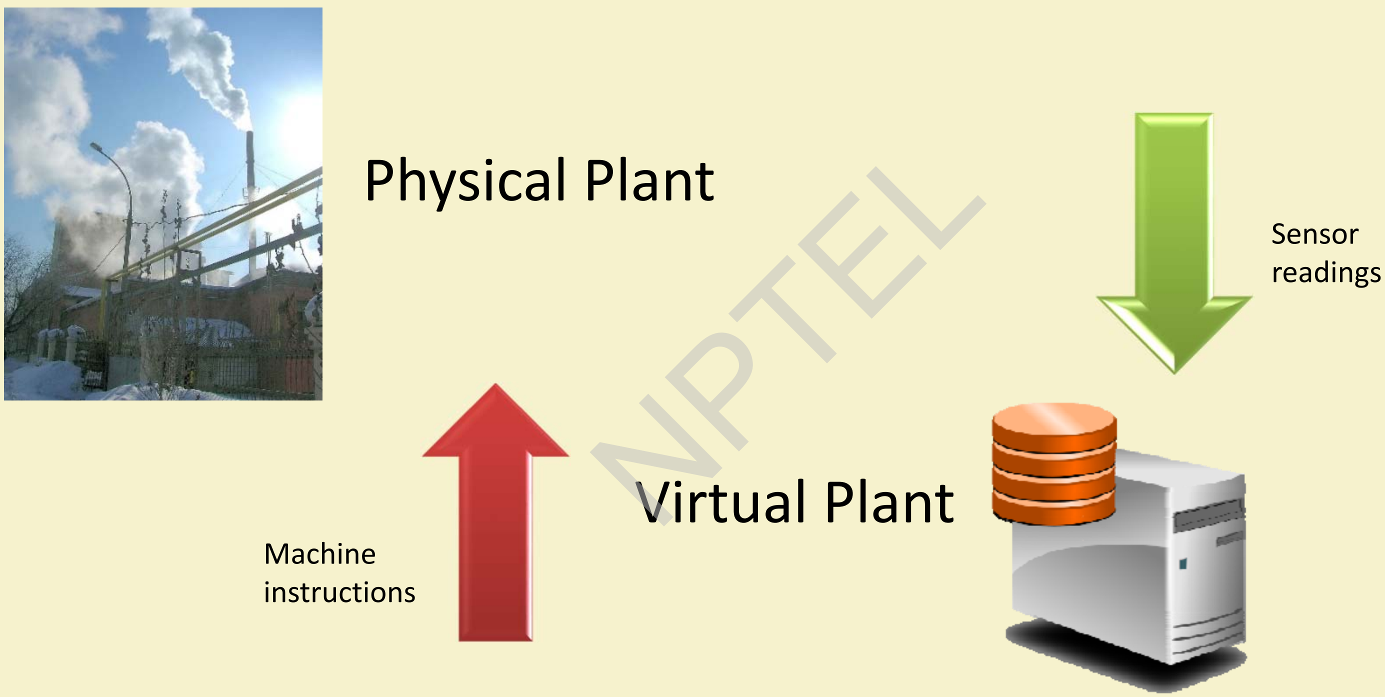
IIoT Requirements



IIoT Requirements (contd.)



IIoT Requirements (contd.)



Design Considerations

- ✓ To use an IoT device for industrial applications, the following design objectives are to be considered –
  - ✓ Energy : Time for which the IoT device can operate with limited power supply.
  - ✓ Latency : Time required to transmit the data.
  - ✓ Throughput : Maximum data transmitted across the network.
  - ✓ Scalability : Number of devices supported.
  - ✓ Topology: Communication among the devices, i.e. interoperability.
  - ✓ Safety and Security: Degree of safety and security of the application.

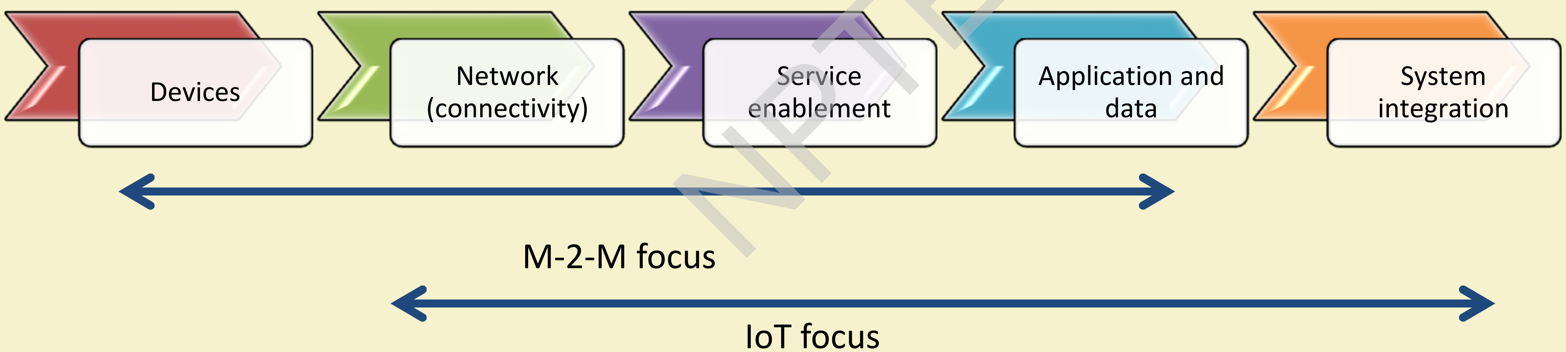
Difference between IoT and IIoT

Difference between IoT and IIoT

The main differences between IoT and IIoT are :

IoT	IIoT
• Focused on convenience of individuals • M-2-M communication: Limited • Applications areas are at consumer-level	• Focused on efficiency, safety and security of the operation. • M-2-M communication: Extensively. • Application areas are at industries.

Difference between IoT and IIoT (contd.)



Service Management in IIoT

- ✓ “Service management refers to the implementation and management of the quality of services which meets the end-users demand”
- ✓ “Service is a collection of data and associated behaviors to accomplish a particular function or feature of a device or portions of a device”.

Source: Ning Lu, Nan Cheng, Ning Zhang, Xuemin Shen, Jon W. Mark, Connected Vehicles : Solutions and Challenges, IEEE Internet of Things Journal, Vol. 1, No. 4, August 2014.

Service Management in IIoT

- ✓ Service can be of two types, which are -
  - ✓ Primary service - The basic services which are responsible for the primary node functions are termed as primary service.
  - ✓ Secondary service - The auxiliary functions which provide services to the primary service or secondary services are termed as secondary service.

Thank You!!

IIT KHARAGPUR

NPTEL ONLINE
CERTIFICATION COURSES

NPTEL

IIoT: Industrial Internet of Things – Part II

Dr. Sudip Misra
Associate Professor
Department of Computer Science and Engineering
IIT Kharagpur
Email: smisra@sit.iitkgp.ernet.in
Website: http://cse.iitkgp.ac.in/~smisra/

Applications of IIoT

- ✓ The key application areas of IIoT are -
  - ✓ Manufacturing industry
  - ✓ Healthcare Service industry
  - ✓ Transportation & logistics
  - ✓ Mining
  - ✓ Firefighting

Manufacturing Industry

- ✓ The devices, equipment, workforce, supply chain, work platform are integrated and connected to achieve smart production. This will led to –
  - ✓ reduction in operational costs
  - ✓ improvement in the productivity of the worker
  - ✓ reduction in the injuries at the workplace
  - ✓ resource optimization and waste reduction
  - ✓ end-to-end automation.

Healthcare Service Industry

- ✓ Patients can be continuously monitored due to the implanted on-body sensors. This has led to –
  - ✓ improved treatment outcome
  - ✓ costs has reduced
  - ✓ improved disease detection
  - ✓ improved accuracy in the collection of data
  - ✓ improved drugs management.

Transportation & logistics

- ✓ To improve safety, efficiency of transportation, Intelligent Transportation system (ITS) is developed which consists of connected vehicles. ITS provides –
  - ✓ Vehicle – to – sensor connectivity
  - ✓ Vehicle – to – vehicle connectivity
  - ✓ Vehicle – to – internet connectivity
  - ✓ Vehicle – to – road infrastructure
- ✓ Dedicated short-range communications (DSRC) is the key enabling technology for V2V and V2R communications.

Transportation & logistics

- ✓ In IIoT scenario the physical objects are provided with
  - ✓ bar codes
  - ✓ RFID tags

  hence, real-time monitoring of the status and location of the physical objects from destination to the origin, across the supply chain is possible.
- ✓ Security and privacy of the data should be maintained.

Mining

- ✓ To prevent accidents inside the mines - RFID, Wi-Fi and other wireless technologies are used, which
  - ✓ provides early warning of any disaster
  - ✓ monitors air-quality
  - ✓ detects the presence of poisonous gases inside the mines
  - ✓ oxygen level inside the mines.

Firefighting

- ✓ Sensor networks, RFID tags are used to perform
  - ✓ automatic diagnosis
  - ✓ early warning of disaster
  - ✓ emergency rescue
  - ✓ provides real-time monitoring

  Hence, improves public security.

Examples of IIoT

- ✓ Examples of IIoT are -
  - ✓ unmanned aerial vehicles (UAVs) to inspect oil pipelines.
  - ✓ monitoring food safety using sensors.
  - ✓ minimizing workers' exposure to noise, chemicals and other hazardous gases.
  - ✓ unmanned marine vehicle which can collect data up to a year without fuel or crew.

Connected Ecosystems in IIoT scenario

- ✓ Traditional supply chains in industries are linear in nature.
- ✓ To shift the business focus from products to outcomes, new ecosystem should be followed.
- ✓ Digital ecosystems progress at a much faster rate than physical industries. Hence, it can quickly adapt to the changes in the external environments.

Integration of Digital and Human Workforce

- ✓ In IIoT, machines become more intelligent. Hence, the automated tasks can be done in the industries at lower costs and higher quality level.
- ✓ Humans will work with machines, the outcome will be higher overall productivity.
- ✓ IIoT will reform and redefine the skills of the workers.

Creation of New Jobs

- ✓ The creation of new composite industries, such as precision agriculture, digital healthcare system, digital mines etc., will lead to development of new job opportunities.
- ✓ Highly automated machines will require lesser number of unskilled workers, but will require skilled experts with digital and analytical skills.

Reformation of Robots

- ✓ In IIoT environment, robots are featured with three capabilities : sensing, thinking and acting. They will be reformed with the ability to carry out repetitive tasks.
- ✓ Robots will be more intelligent but will work under the supervision of human beings. Their availability will increase.
- ✓ Robots will be reprogrammable to perform new tasks. They have the capability to 'learn' faster.

Challenges in IIoT

- ✓ Primary challenges

Identification of objects or things

Manage huge amount of data

Integrate existing infrastructures into new IIoT infrastructure

Enabling data storage

Challenges in IIoT(contd.)

✓ Safety Challenges

- Worker health and safety
- Regulatory compliance
- Environmental protection
- Optimized operations

Challenges in IIoT(contd.)

✓ Hazards (related)

- Handling, storing or using hazardous substances
- Oxygen deficiency
- Particulates
- Radiation
- Physiological stress

Challenges in IIoT(contd.)

<u>Standardization</u>

- ✓ Standardization plays an important role in the development of the system.
- ✓ Goal: To improve the interoperability of the different systems/ applications and allow the products/services to perform better.

Challenges in IIoT(contd.)

Standardization

- ✓ The problems related to standardization are:
  - ✓ Interoperability
  - ✓ Semantic interoperability (data sematics)
  - ✓ Security and privacy
  - ✓ Radio access level issues.

Challenges in IIoT(contd.)

<u>Privacy and security issues</u>

- ✓ The two most important concerns related with IIoT are -
  - ✓ information security
  - ✓ data privacy protection
- ✓ The devices/things can be tracked, monitored and connected. So there are chances of attack on the personal and private data.

Challenges in IIoT(contd.)

Privacy and security issues

- ✓ Examples –
  - ✓ Healthcare industry – the medical data of a patient must not be tampered, or altered by any person in the middle.
  - ✓ Food industry – the deterioration of any food item being sent to the company must be kept confidential as it will affect the reputation of the company.

Risks associated with IIoT in Manufacturing

- ✓ Though IIoT provides new opportunities, but few factors may cause hindrance in the path to success, which are :
  - ✓ lack of vision and leadership
  - ✓ lack of understanding of values among management employees
  - ✓ costly sensors
  - ✓ inadequate infrastructure.

Meet the challenges: Sensor improvement

- ✓ Improvement in sensor technologies –
  - ✓ miniaturization
  - ✓ performance
  - ✓ cost and energy consumption.

Meet the challenges : Manufacturing

- ✓ Manufacturers use software capabilities to improve operational efficiency through –
  - ✓ predictive maintenance
  - ✓ savings on scheduled repairs
  - ✓ reduced maintenance costs
  - ✓ reduced number of breakdowns.

Case study : Rt Tech Software

- ✓ Rt Tech particularizes in software which –
  - ✓ improves industrial facilities' efficiency
  - ✓ improves productivity.
- ✓ Energy management solution, which leads to reduction in the plant's highest variable cost.
- ✓ Rt Tech automates the process of mapping and managing energy consumption.

Source : http://www.mcrockcapital.com

PRODUCTS DEVELOPED

- ✓ M-2-M communication : Intelligent Radio Modem (IRM)
  - ✓ IRM 1500 & ACE 1000 - IRM
    - ✓ simple
    - ✓ M-2-M connectivity
    - ✓ data transmission
  - ✓ These devices provide easy maintenance and installation. They can be connected to IP and non-IP serial devices to extend the capability to monitor and communicate with other technologies.

Source : https://www.motorolasolutions.com

PRODUCTS DEVELOPED (contd.)

- ✓ Comtrol – IO Link Master Gateway
  - ✓ It can be easily integrated into the industrial network with existing and new installations.
  - ✓ It supports Ethernet/IP, PROFINET (PNIO) and Modbus TCP.

Source :
http://pdfserv.maximintegrated.com
http://www.comtrol.com

Benefits of IIoT

✓ The benefits of IIoT are

- Improved connectivity among devices
- Improved efficiency
- Upgraded scalability
- Reduces operation time
- Remote diagnosis
- Cost effective

Recent Research trends in IIoT

- ✓ Recent research challenges in IIoT are -
  - ✓ To improve the communications among the different things or objects.
  - ✓ To develop energy-efficient techniques so as to reduce power consumption by sensors.
  - ✓ To develop context-aware IoT middleware for better understanding of the sensor data.
  - ✓ To create smart objects with larger memory, processing and reasoning capabilities.

Conclusion

- ✓ IIoT system requires the following :
  - ✓ Smaller, less expensive sensors which makes them easily accessible.
  - ✓ Distributed control of assembly line, automated monitoring, control and maintenance.

References

- ✓ Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, Imrich Chlamtac, Internet of things: Vision, applications and research challenges, Ad Hoc Networks, Volume 10, Issue 7, September 2012.
- ✓ http://internetofthingsagenda.techtarget.com/definition/Industrial-Internet-of-Things-IIoT.
- ✓ Ning Lu, Nan Cheng, Ning Zhang, Xuemin Shen, Jon W. Mark, Connected Vehicles : Solutions and Challenges, IEEE Internet of Things Journal, Vol. 1, No. 4, August 2014.
- ✓ Zhibo Pang, Qiang Chen, Junzhe Tian, Lirong Zheng and E. Dubrova, Ecosystem analysis in the design of open platform-based in-home healthcare terminals towards the internet-of-things, 2013, 15th International Conference on Advanced Communications Technology (ICACT), PyeongChang, 2013.
- ✓ Wei Qiuping, Zhu Shunbing, Du Chunquan, Study On Key Technologies Of Internet Of Things Perceiving Mine, Procedia Engineering, Volume 26, 2011.
- ✓ Bill Karakostas, A DNS Architecture for the Internet of Things: A Case Study in Transport Logistics, Procedia Computer Science, Volume 19, 2013.
- ✓ Ying-cong Zhang, Jing Yu, A Study on the Fire IOT Development Strategy, Procedia Engineering, Volume 52, 2013.

References

- ✓ J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, Internet of things(IoT): A vision, architectural elements, and future directions, Future Gen. Comput. Syst., vol. 29, no. 7, 2013 .
- ✓ D. Bandyopadhyay and Jaydip Sen, Internet of things: Applications and challenges in technology and standardization, Wireless Personal Communications 58.1 (2011).
- ✓ Industry 4.0, The Industrial Internet of Things, by Alasdair Gilchrist
- ✓ http://pdfserv.maximintegrated.com
- ✓ http://www.comtrol.com
- ✓ http://www.mcrockcapital.com
- ✓ http://web.stanford.edu
- ✓ http://www.accenture.com

Thank You!!

Data Handling and Analytics – Part I

Data is Precious

Dr. Sudip Misra
Associate Professor
Department of Computer Science and Engineering
IIT KHARAGPUR
Email: smisra@sit.iitkgp.ernet.in
Website: http://cse.iitkgp.ac.in/~smisra/

What is Data Handling

- ✓ Data handling
  - ✓ Ensures that research data is stored, archived or disposed off in a safe and secure manner during and after the conclusion of a research project
  - ✓ Includes the development of policies and procedures to manage data handled electronically as well as through non-electronic means.
- ✓ In recent days, most data concern –
  - ✓ Big Data
    - ✓ Due to heavy traffic generated by IoT devices
    - ✓ Huge amount of data generated by the deployed sensors

What is Big Data

- ✓ *"Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling the high-velocity capture, discovery, and/or analysis."*

  [Report of International Data Corporation (IDC)]

- ✓ *"Big data shall mean the data of which the data volume, acquisition speed, or data representation limits the capacity of using traditional relational methods to conduct effective analysis or the data which may be effectively processed with important horizontal zoom technologies."*

  [National Institute of Standards and Technology (NIST)]

Types of Data

- ✓ Structured data
  - ✓ Data that can be easily organized.
  - ✓ Usually stored in relational databases.
  - ✓ Structured Query Language (SQL) manages structured data in databases.
  - ✓ It accounts for only 20% of the total available data today in the world.
- ✓ Unstructured data
  - ✓ Information that do not possess any pre-defined model.
  - ✓ Traditional RDBMSs are unable to process unstructured data.
  - ✓ Enhances the ability to provide better insight to huge datasets.
  - ✓ It accounts for 80% of the total data available today in the world.

Characteristics of Big Data

- ✓ Big Data is characterized by 7 Vs –
  - ✓ **V**olume
  - ✓ **V**elocity
  - ✓ **V**ariety
  - ✓ **V**ariability
  - ✓ **V**eracity
  - ✓ **V**isualization
  - ✓ **V**alue

Characteristics of Big Data (Contd.)

- ✓ Volume
  - ✓ Quantity of data that is generated
  - ✓ Sources of data are added continuously
  - ✓ Example of *volume* -
    - ✓ 30TB of images will be generated every night from the Large Synoptic Survey Telescope (LSST)
    - ✓ 72 hours of video are uploaded to YouTube every minute

Characteristics of Big Data (Contd.)

- Velocity
  - Refers to the speed of generation of data
  - Data processing time decreasing day-by-day in order to provide real-time services
  - Older batch processing technology is unable to handle high velocity of data
  - Example of *velocity* –
    - 140 million tweets per day on average (according to a survey conducted in 2011)
    - New York Stock Exchange captures 1TB of trade information during each trading session

Characteristics of Big Data (Contd.)

- ✓ Variety
  - ✓ Refers to the category to which the data belongs
  - ✓ No restriction over the input data formats
  - ✓ Data mostly unstructured or semi-structured
  - ✓ Example of *variety* –
    - ✓ Pure text, images, audio, video, web, GPS data, sensor data, SMS, documents, PDFs, flash etc.

Characteristics of Big Data (Contd.)

- ✓ Variability
  - ✓ Refers to data whose meaning is constantly changing.
  - ✓ Meaning of the data depends on the context.
  - ✓ Data appear as an indecipherable mass without structure
  - ✓ Example:
    - ✓ Language processing, Hashtags, Geo-spatial data, Multimedia, Sensor events
- ✓ Veracity
  - ✓ Veracity refers to the biases, noise and abnormality in data.
  - ✓ It is important in programs that involve automated decision-making, or feeding the data into an unsupervised machine learning algorithm.
  - ✓ Veracity isn't just about data quality, it's about data understandability.

Characteristics of Big Data (Contd.)

- ✓ Visualization
  - ✓ Presentation of data in a pictorial or graphical format
  - ✓ Enables decision makers to see analytics presented visually
  - ✓ Identify new patterns

- ✓ Value
  - ✓ It means extracting useful business information from scattered data.
  - ✓ Includes a large volume and variety of data
  - ✓ Easy to access and delivers quality analytics that enables informed decisions

Data Handling Technologies

- ✓ Cloud computing
  - ✓ Essential characteristics according to NIST
    - ✓ On-demand self service
    - ✓ Broad network access
    - ✓ Resource pooling
    - ✓ Rapid elasticity
    - ✓ Measured service
  - ✓ Basic service models provided by cloud computing
    - ✓ Infrastructure-as-a-Service (IaaS)
    - ✓ Platform-as-a-Service (PaaS)
    - ✓ Software-as-a-Service (SaaS)

Data Handling Technologies (Contd.)

- ✓ Internet of Things (IoT)
  - ✓ According to Techopedia, IoT *"describes a future where every day physical objects will be connected to the internet and will be able to identify themselves to other devices."*
  - ✓ Sensors embedded into various devices and machines and deployed into fields.
  - ✓ Sensors transmit sensed data to remote servers via Internet.
  - ✓ Continuous data acquisition from mobile equipment, transportation facilities, public facilities, and home appliances

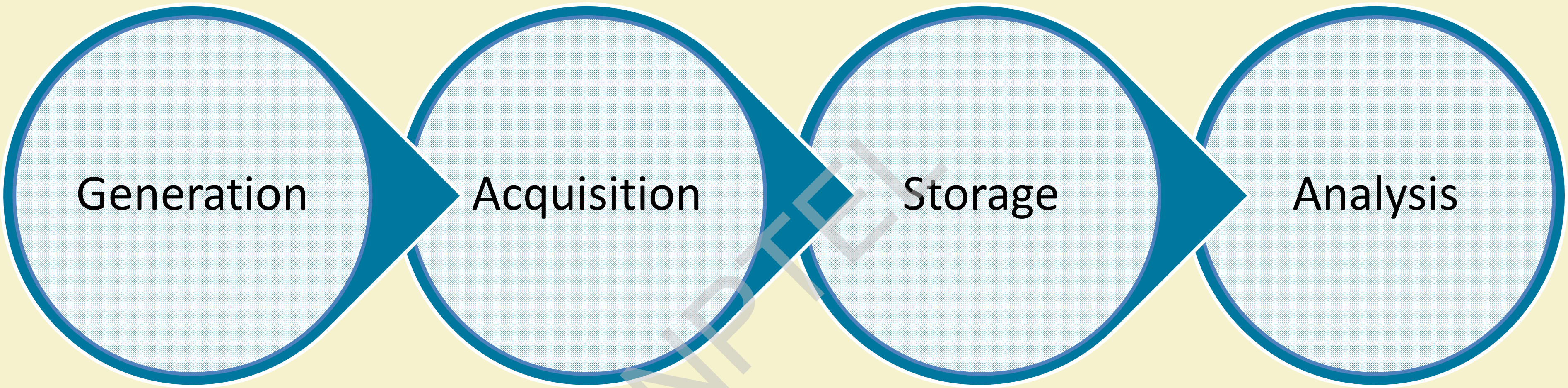
Data Handling Technologies (Contd.)

- ✓ Internet of Things (IoT)
  - ✓ According to Techopedia, IoT *“describes a future where every day physical objects will be connected to the internet and will be able to identify themselves to other devices.”*
  - ✓ Sensors embedded into various devices and machines and deployed into fields.
  - ✓ Sensors transmit sensed data to remote servers via Internet.
  - ✓ Continuous data acquisition from mobile equipment, transportation facilities, public facilities, and home appliances

Data Handling Technologies (Contd.)

- ✓ Data handling at data centers
  - ✓ Storing, managing, and organizing data.
  - ✓ Estimates and provides necessary processing capacity.
  - ✓ Provides sufficient network infrastructure.
  - ✓ Effectively manages energy consumption.
  - ✓ Replicates data to keep backup.
  - ✓ Develop business oriented strategic solutions from big data.
  - ✓ Helps business personnel to analyze existing data.
  - ✓ Discovers problems in business operations.

Flow of Data



- ✓ Enterprise data
- ✓ IoT data
- ✓ Bio-medical data
- ✓ Other data

- ✓ Data collection
- ✓ Data transportation
- ✓ Data pre-processing

- ✓ Hadoop
- ✓ MapReduce
- ✓ NoSQL databases

- ✓ Bloom filter
- ✓ Parallel computing
- ✓ Hashing and indexing

Data Sources

- ✓ Enterprise data
  - ✓ Online trading and analysis data.
  - ✓ Production and inventory data.
  - ✓ Sales and other financial data.
- ✓ IoT data
  - ✓ Data from industry, agriculture, traffic, transportation
  - ✓ Medical-care data,
  - ✓ Data from public departments, and families.
- ✓ Bio-medical data
  - ✓ Masses of data generated by gene sequencing.
  - ✓ Data from medical clinics and medical R&Ds.
- ✓ Other fields
  - ✓ Fields such as – computational biology, astronomy, nuclear research etc

Data Acquisition

- ✓ Data collection
  - ✓ Log files or record files that are automatically generated by data sources to record activities for further analysis.
  - ✓ Sensory data such as sound wave, voice, vibration, automobile, chemical, current, weather, pressure, temperature etc.
  - ✓ Complex and variety of data collection through mobile devices. E.g. – geographical location, 2D barcodes, pictures, videos etc.
- ✓ Data transmission
  - ✓ After collecting data, it will be transferred to storage system for further processing and analysis of the data.
  - ✓ Data transmission can be categorized as – Inter-DCN transmission and Intra-DCN transmission.

Data Acquisition (Contd.)

- ✓ Data pre-processing
  - ✓ Collected datasets suffer from noise, redundancy, inconsistency etc., thus, pre-processing of data is necessary.
  - ✓ Pre-processing of relational data mainly follows – integration, cleaning, and redundancy mitigation
  - ✓ Integration is combining data from various sources and provides users with a uniform view of data.
  - ✓ Cleaning is identifying inaccurate, incomplete, or unreasonable data, and then modifying or deleting such data.
  - ✓ Redundancy mitigation is eliminating data repetition through detection, filtering and compression of data to avoid unnecessary transmission.

Data Storage

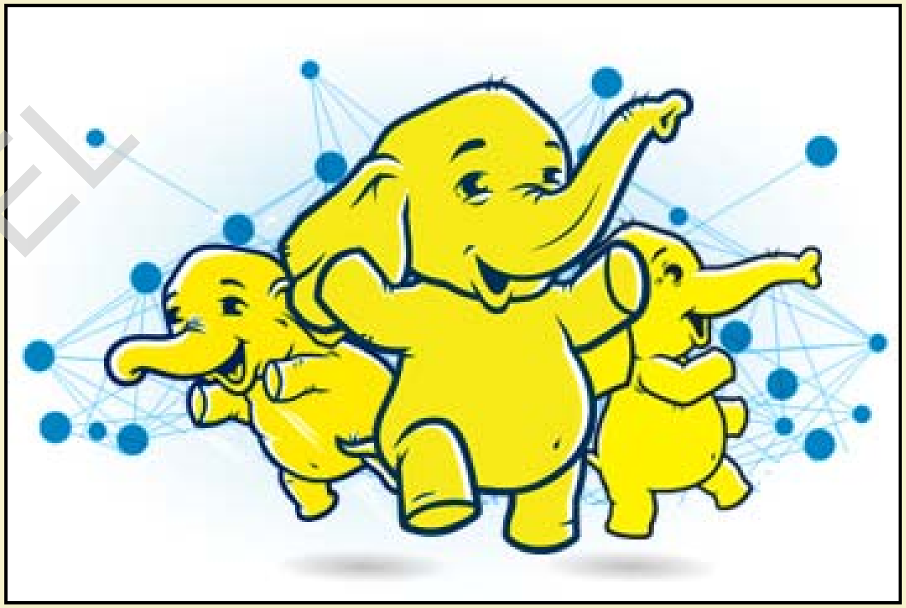
- ✓ File system
  - ✓ Distributed file systems that store massive data and ensure – consistency, availability, and fault tolerance of data.
  - ✓ GFS is a notable example of distributed file system that supports large-scale file system, though it's performance is limited in case of small files
  - ✓ Hadoop Distributed File System (HDFS) and Kosmosfs are other notable file systems, derived from the open source codes of GFS.
- ✓ Databases
  - ✓ Emergence of non-traditional relational databases (NoSQL) in order to deal with the characteristics that big data possess.
  - ✓ Three main NoSQL databases – Key-value databases, column-oriented databases, and document-oriented databases.

Data Handling Using Hadoop

Reliable, scalable, distributed data handling

What is Hadoop

- ✓ Hadoop is a software framework for distributed processing of large datasets across large clusters of computers.
- ✓ Hadoop is open-source implementation for Google 's GFS and ***MapReduce.***
- ✓ Apache Hadoop's Map Reduce and Hadoop Distributed File System (HDFS) components originally derived respectively from Google's MapReduce and Google File System (GFS) .



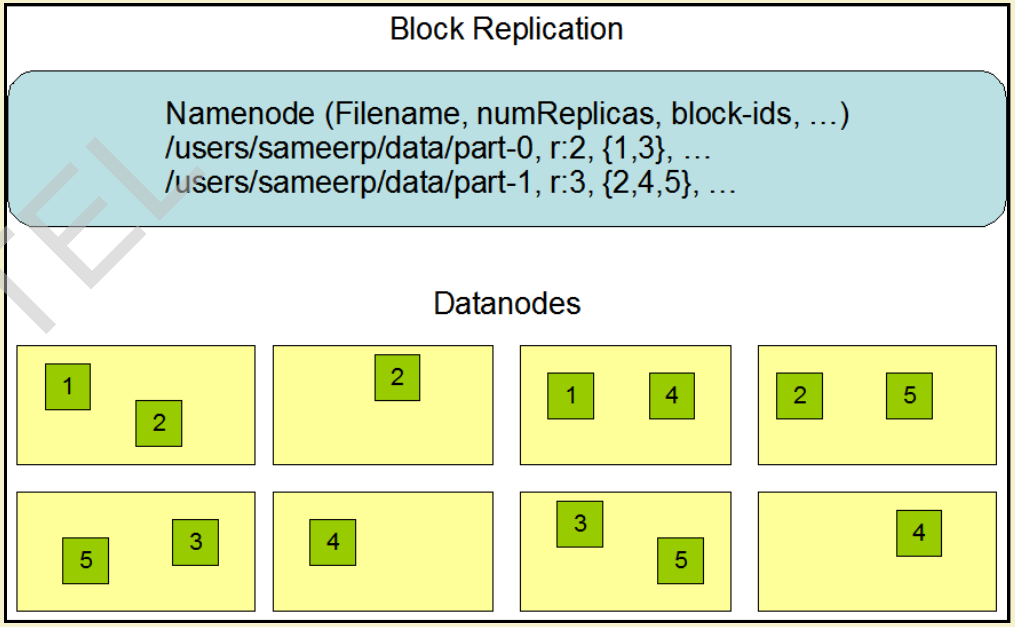
Source: https://www.cloudnloud.com/hadoop-hdfs-operations/

Building Blocks of Hadoop

- ✓ Hadoop Common
  - ✓ A module containing the utilities that support the other Hadoop components
- ✓ Hadoop Distributed File System (HDFS)
  - ✓ Provides reliable data storage and access across the nodes
- ✓ MapReduce
  - ✓ Framework for applications that process large amount of datasets in parallel.
- ✓ Yet Another Resource Negotiator (YARN)
  - ✓ Next-generation MapReduce, which assigns CPU, memory and storage to applications running on a Hadoop cluster.

Hadoop Distributed File System (HDFS)

- ✓ Centralized node
  - ✓ Namenode
  - ✓ Maintains metadata info about files
- ✓ Distributed node
  - ✓ Datanode
  - ✓ Store the actual data
  - ✓ Files are divided into blocks
  - ✓ Each block is replicated



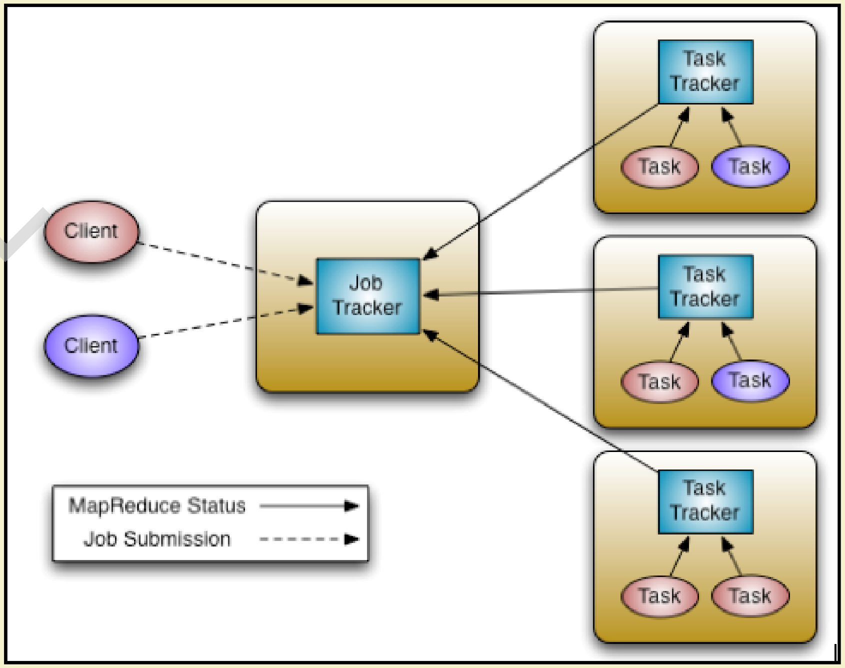
Source: http://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

Name and Data Nodes

- ✓ Namenode
  - ✓ Stores filesystem metadata.
  - ✓ Maintains two in-memory tables, to map the datanodes to the blocks, and vice versa
- ✓ Datanode
  - ✓ Stores actual data
  - ✓ Data nodes can talk to each other to rebalance and replicate data
  - ✓ Data nodes update the namenode with the block information periodically
  - ✓ Before updating datanodes verify the checksums.

Job and Task Trackers

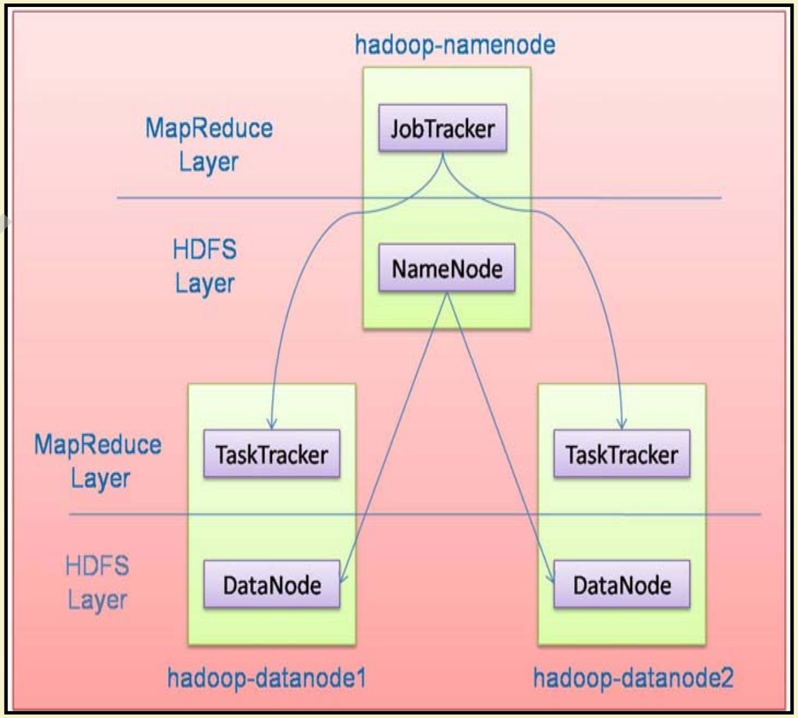
- ✓ Job Tracker –
  - ✓ Runs with the Namenode
  - ✓ Receives the user's job
  - ✓ Decides on how many tasks will run (number of mappers)
  - ✓ Decides on where to run each mapper (concept of locality)
- ✓ Task Tracker –
  - ✓ Runs on each datanode
  - ✓ Receives the task from Job Tracker
  - ✓ Always in communication with the Job Tracker reporting progress



Source: http://developeriq.in/articles/2015/aug/11/an-introduction-to-apache-hadoop-for-big-data/

Hadoop Master/Slave Architecture

- ✓ *Master-slave shared-nothing* architecture
- ✓ Master
  - ✓ Executes operations like opening, closing, and renaming files and directories.
  - ✓ Determines the mapping of blocks to Datanodes.
- ✓ Slave
  - ✓ Serves read and write requests from the file system's clients.
  - ✓ Performs block creation, deletion, and replication as instructed by the Namenode.



Source: http://ankitasblogger.blogspot.in/2011/01/hadoop-cluster-setup.html

References

- ✓ R. Ahmed and G. Karypis, “Algorithms for Mining the Evolution of Conserved Relational States in Dynamic Networks,” Knowledge and Information Systems, vol. 33, no. 3, pp. 603-630, Dec. 2012.
- ✓ M.H. Alam, J.W. Ha, and S.K. Lee, “Novel Approaches to Crawling Important Pages Early,” Knowledge and Information Systems, vol. 33, no. 3, pp 707-734, Dec. 2012.
- ✓ S. Aral and D. Walker, “Identifying Influential and Susceptible Members of Social Networks,” Science, vol. 337, pp. 337-341, 2012.
- ✓ A. Machanavajjhala and J.P. Reiter, “Big Privacy: Protecting Confidentiality in Big Data,” ACM Crossroads, vol. 19, no. 1, pp. 20-23, 2012.
- ✓ S. Banerjee and N. Agarwal, “Analyzing Collective Behavior from Blogs Using Swarm Intelligence,” Knowledge and Information Systems, vol. 33, no. 3, pp. 523-547, Dec. 2012.
- ✓ E. Birney, “The Making of ENCODE: Lessons for Big-Data Projects,” Nature, vol. 489, pp. 49-51, 2012.
- ✓ S. Borgatti, A. Mehra, D. Brass, and G. Labianca, “Network Analysis in the Social Sciences,” Science, vol. 323, pp. 892-895, 2009.
- ✓ J. Bughin, M. Chui, and J. Manyika, Clouds, Big Data, and Smart Assets: Ten Tech-Enabled Business Trends to Watch. McKinSey Quarterly, 2010.
- ✓ D. Centola, “The Spread of Behavior in an Online Social Network Experiment,” Science, vol. 329, pp. 1194-1197, 2010.
- ✓ http://hadoop.apache.org/

Thank You!!